

Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo Methods

Saïd Moussaoui, David Brie, Cédric Carteret, Ali Mohammad-Djafari

► **To cite this version:**

Saïd Moussaoui, David Brie, Cédric Carteret, Ali Mohammad-Djafari. Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo Methods. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, 2006, 81(2), pp.137-148. 10.1016/j.chemolab.2005.11.004 . hal-00022304

HAL Id: hal-00022304

<https://hal.archives-ouvertes.fr/hal-00022304>

Submitted on 6 Apr 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian Analysis of Spectral Mixture Data using Markov Chain Monte Carlo Methods

Saïd Moussaoui^{a 1}, Cédric Carteret^b, David Brie^a, Ali Mohammad-Djafari^c

^a *Université Henri Poincaré, Nancy 1, Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039 CNRS-UHP-INPL, B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France.*

^b *Université Henri Poincaré, Nancy 1, Laboratoire de Chimie Physique et Microbiologie pour l'Environnement (LCPME), UMR 7564 CNRS-UHP, 54600 Villers-lès-Nancy, France.*

^c *Laboratoire des Signaux et Systèmes (LSS), CNRS-SUPÉLEC-UPS, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France.*

Abstract

This paper presents an original method for the analysis of multicomponent spectral data sets. The proposed algorithm is based on Bayesian estimation theory and Markov Chain Monte Carlo (MCMC) methods. Resolving spectral mixture analysis aims at recovering the unknown component spectra and at assessing the concentrations of the underlying species in the mixtures. In addition to non-negativity constraint, further assumptions are generally needed to get a unique resolution. The proposed statistical approach assumes mutually independent spectra and accounts for the non-negativity and the sparsity of both the pure component spectra and the concentration profiles. Gamma distribution priors are used to translate all these information in a probabilistic framework. The estimation is performed using MCMC methods which lead to an unsupervised algorithm, whose performances are assessed in a simulation study with a synthetic data set.

Key words: Curve Resolution, Factor analysis, Bayesian estimation, non-negativity, statistical independence, sparsity, Gamma distribution, Markov Chain Monte Carlo (MCMC).

Journal article

Chemometrics and Intelligent Laboratory Systems,
volume 81, issue 2, pages: 137-148, April 2006

¹ Corresponding author: Saïd Moussaoui. Tel: (+33) 383 684 495;
Fax: (+33) 383 684 462; E-mail: said.moussaoui@cran.uhp-nancy.fr

1 Introduction

Spectral data sets resulting from observations of multicomponent substances are interpreted as a weighted sum of the unknown pure component spectra. The mixing model assumes that m measured data $\{D_{(i,k)}, k = 1, \dots, n\}_{i=1}^m$ are linear combinations of p unknown pure component spectra $\{S_{(j,k)}, k = 1, \dots, n\}_{j=1}^p$. Mathematically, this model is expressed as

$$D_{(i,k)} = \sum_{j=1}^p C_{(i,j)} S_{(j,k)} + E_{(i,k)}, \quad (1)$$

where $i = 1, \dots, m$ and $j = 1, \dots, p$, respectively, index the measured spectra and the unknown pure component spectra and the index k corresponds to the spectral variable $\{\lambda_k, k = 1, \dots, n\}$. Each mixing coefficient $C_{(i,j)}$ is proportional to the concentration of the j -th pure component in the i -th mixture. The additive noise terms $\{E_{(i,k)}, k = 1, \dots, n\}_{i=1}^m$ represent the measurement errors and model imperfections. Using matrix notations this model is written as

$$\mathbf{D} = \mathbf{C} \mathbf{S} + \mathbf{E}, \quad (2)$$

where the row vectors of the $(m \times n)$ data matrix \mathbf{D} contain the m measured spectra, \mathbf{C} is the $(m \times p)$ mixing matrix, with its column vectors representing the mixing coefficient profiles of the pure components. \mathbf{S} is the $(p \times n)$ matrix, with its row vectors containing the p pure component spectra and \mathbf{E} is the $(m \times n)$ noise matrix.

By assuming a known number of components, the mixture analysis aim is to jointly estimate the pure component spectra and the mixing coefficient profiles. For a qualitative analysis, pure spectra are necessary to identify the chemical composition of the substances and the mixing coefficients are used, in a quantitative analysis, to evaluate the concentrations of each component. In the chemometrics community, this problem is termed by *curve resolution* [1], *mixture analysis* [2] and *factor analysis* [3] while in the signal processing field this problem is called *blind source separation* [4]. In the sequel, the problem is referred to as spectral mixture analysis.

This paper proposes a statistical mixture analysis method using Bayesian estimation theory, which consists in assigning prior distributions to account for any available knowledge on pure spectra and mixing coefficients. The probabilistic formulation of the problem enables the use of stochastic simulation tools to draw samples from the posterior distribution from which estimates of means, variances and other statistical characteristics can be computed. In order to state clearly the contributions of this paper, section 2 gives a short overview on existing spectral mixture analysis methods which are mainly presented regarding the underlying assumptions

and resolution algorithms, thus allowing to point out some of their limitations that may be handled by the Bayesian/MCMC approach. The probabilistic formulation of the problem is detailed and a discussion on its relation with the *positive matrix factorization* (PMF) method is given. The method is illustrated in section 4 with a synthetic data set and a discussion of the behavior of the method.

2 An Overview on Spectral Mixture Analysis Methods

In this section a short overview of spectral mixture analysis methods is given. A more complete state of the art may be found in [5, 6]. A mixture analysis algorithm is defined firstly, by the assumptions made on the pure component spectra and their concentration profiles which lead to an objective function, and secondly by the optimization algorithm used to optimize this objective function. The main constraint in spectral mixture data analysis is the non-negativity of both the pure component spectra and concentrations. However, only accounting for this constraint does not lead to unique solution [1]. Thus, additional constraints or assumptions are required to select a particular solution. According to the way these constraints and assumptions are introduced, the available mixture analysis methods can be classified into two main categories. The first one, termed as algebraic methods, is based on Lawton-Sylvestre approach [1], termed by *multivariate curve resolution*, while the second is based on a *non-negative least squares estimation* [7, 8].

2.1 Algebraic Methods

As illustrated by figure 1, this approach firstly decomposes the data matrix \mathbf{D} , using principal component analysis (PCA), and then finds a linear transformation \mathbf{T} that transforms the principal components \mathbf{V} and their weight matrix \mathbf{U} into non-negative estimates of pure spectra and mixing coefficients.

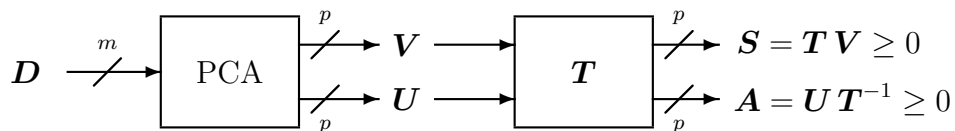


Figure 1. Lawton-Sylvestre Approach

Since only accounting for non-negativity does not ensure the uniqueness of the solution, this

approach leads to a set of admissible solutions [1, 9–12]. In order to reduce this set [13, 14] suggest to add further constraints, in addition to the non-negativity, and propose to search a linear transformation by minimizing a compound criterion, in which a first part penalizes negative estimates of the pure spectra and mixing coefficients, and the second part uses an entropic cost function to make the estimated spectra smoother and mutually independent. The optimization of the whole objective function is performed by using the simplex method. However, in spite of the theoretical uniqueness of the solution by considering the additional *a priori* constraints, this method may converge to local or false minimums. This is due to the fact that the convexity of the criterion is not proven and the global criterion shape is highly depending on the regularization parameters that are specified manually. Recently, this method have been revisited in [15, 16] where a simulated annealing optimization algorithm has been used, but this algorithm may also converge to a local minimum. In the signal processing community, a similar approach to Lawton-Sylvestre method, based on statistical independence of pure spectra and called *independent component analysis* [4, 17–19], is used for the analysis of mixtures. To get the mutually independent signals, the transformation matrix is reduced to a unitary rotation matrix. Assuming the statistical independence, ensures the uniqueness of the solution, but it does not guarantee its non-negativity. To handle this problem, the *non-negative independent component analysis* (NNICA) [20], finds a transformation matrix by optimizing an objective function that enforces the non-negativity of the estimated pure spectra. However, this method assumes orthogonal pure spectra and does not account explicitly the mixing coefficient non-negativity.

2.2 Non-negative Least Squares Estimation

This approach performs constrained least squares estimation and there are mainly two methods: *alternating least squares* (ALS) [7, 8] and *non-negative matrix factorization* (NMF) [21]. ALS performs a decomposition of the mixture by minimizing at each iteration the least squares criterion under the non-negativity constraint [22, 23]. The NMF method minimizes the least squares criterion by using a gradient descent algorithm over this objective function and updates iteratively concentrations and pure spectra using a particular multiplicative learning rule that ensures the estimates to remain non-negative. The key point is that the solution using only non-negativity is not unique, therefore the results provided by ALS and NMF methods depend on their initializations. In practice, the NMF method is randomly initialized [24], while the ALS method is initialized by the results obtained with a non-constrained decomposition method such as principal component analysis (PCA), factor analysis algorithms [25–27], or using pure variable detection methods such as *simple-to-use interactive self modeling mixture*

analysis (SIMPLISMA) [28] and orthogonal projection approach (OPA) [29]. More recently ICA methods are also used [30, 31] as an initialization method. Similarly to the algebraic methods, additional constraints such as closure, unimodality, selectivity may be added to reduce the set of admissible solutions [32–34]. These additional constraints are accounted through either a penalized least squares estimation or a constrained estimation. In the penalized least square methods, a regularization criterion is added to the weighted mean squares criterion in order to select a particular solution fulfilling some assumptions defined *a priori*. This is the basis of methods such as *positive matrix factorization* (PMF) [35, 36] or *non-negative sparse coding* (NNSC) [37]. Unlike the constrained least squares methods, the PMF approach leads to an unconstrained optimization problem and ensures the uniqueness of the solution for a fixed set of regularization parameters. Coming from the fact that any penalized least squares estimation method can be formulated in a Bayesian framework, we propose a fully Bayesian approach to address the inverse problem of spectral mixture analysis.

3 A Bayesian Approach to Spectral Mixture Analysis

Spectral mixture analysis is an ill-posed inverse problem since the solution is not unique. Therefore, to get a preferable solution one has to use any available knowledge or additional assumptions. Therefore, this inverse problem can be successfully addressed using Bayesian estimation theory by encoding the available information through the assignment of *a priori* distributions on the pure component spectra and concentration profiles. The formulation of source separation and factor analysis using Bayesian estimation theory has been suggested recently in [38–41] and used in signal and image processing problems. However, its application to the separation of spectral mixture data has only received a few attention [42–44]. Since only accounting for the non-negativity, mixture analysis still remains ill-posed, one has to use additional constraints or assumptions on the pure spectra and concentrations to get a unique solution. In this paper, we introduce an assumption on their statistical distributions which aims at encoding the non-negativity and the pure spectra sparsity. Before going further, let us recall the Bayesian approach to mixture analysis.

3.1 Bayesian Approach

The main idea of the Bayesian approach to mixture analysis is to encode the available knowledge on the pure spectra and concentration profiles through the assignment of prior distributions $p(\mathbf{S})$ and $p(\mathbf{C})$, which according to Bayes’ theorem and using the likelihood $p(\mathbf{D}|\mathbf{S}, \mathbf{C})$

leads to the *a posteriori* density

$$p(\mathbf{S}, \mathbf{C} | \mathbf{D}) = p(\mathbf{D} | \mathbf{S}, \mathbf{C}) \times p(\mathbf{S}) \times p(\mathbf{C}) \div p(\mathbf{D}), \quad (3)$$

where the independence between \mathbf{C} and \mathbf{S} is assumed. Since $p(\mathbf{D})$ is a normalization constant, one can write

$$p(\mathbf{S}, \mathbf{C} | \mathbf{D}) \propto p(\mathbf{D} | \mathbf{S}, \mathbf{C}) \times p(\mathbf{S}) \times p(\mathbf{C}). \quad (4)$$

This *a posteriori* density combines explicitly the *a priori* knowledge and assumptions on the pure component spectra and concentrations with the information coming from the measured data.

To get estimates of pure component spectra and concentrations, various Bayesian estimators [45] can be applied

- the *joint maximum a posteriori* (JMAP) estimates are obtained by

$$\left(\hat{\mathbf{S}}^{JMAP}, \hat{\mathbf{C}}^{JMAP} \right) = \arg \max_{\mathbf{S}, \mathbf{C}} p(\mathbf{S}, \mathbf{C} | \mathbf{D}). \quad (5)$$

This maximisation is equivalent to a joint minimisation of the corresponding criterion, defined by

$$J(\mathbf{S}, \mathbf{C}) = -\log p(\mathbf{S}, \mathbf{C} | \mathbf{D}), \quad (6)$$

and depends not only on the measured data but also on the prior distributions of pure spectra and concentration profiles

$$J(\mathbf{S}, \mathbf{C}) = -\log p(\mathbf{D} | \mathbf{S}, \mathbf{C}) - \log p(\mathbf{S}) - \log p(\mathbf{C}). \quad (7)$$

Each prior model leads to a particular regularization criterion. This aspect can be seen as an implicit penalty function synthesis through a probabilistic formulation of the prior knowledge. This formulation is known as a Bayesian interpretation of penalized least squares estimation methods.

- the *marginal maximum a posteriori* (MMAP) estimates are obtained by integrating (marginalizing) with respect to \mathbf{S} or \mathbf{C} and maximizing the resulting posterior marginal distribution $p(\mathbf{C} | \mathbf{D})$ or $p(\mathbf{S} | \mathbf{D})$.
- the *marginal posterior mean* (MPM) estimates are obtained from the mean of the marginal posterior distributions $p(\mathbf{C} | \mathbf{D})$ and $p(\mathbf{S} | \mathbf{D})$. In practice, the marginalization is generally performed using MCMC methods and the posterior means are calculated from the simulated Markov chains.

In the following, the proposed Bayesian inference for the analysis of spectral mixtures and its related estimation algorithm are presented. The mathematical computation details can be

found in [44].

3.2 Noise Distribution and Likelihood

Each noise sequence $\{E_{(i,k)}, k = 1, \dots, n\}$ is assumed Gaussian, zero mean, independent and identically distributed with variance σ_i^2 . Moreover, all the noise sequences are assumed mutually independent. Thus,

$$p(\mathbf{E}|\boldsymbol{\theta}_1) = \prod_{i=1}^m \prod_{k=1}^n \mathcal{N}(E_{(i,k)}; 0, \sigma_i^2), \quad (8)$$

where $\boldsymbol{\theta}_1 = [\sigma_1^2, \dots, \sigma_m^2]^T$ and $\mathcal{N}(x; \mu, \nu)$ denotes the normal distribution for the variable x with mean μ and variance ν

$$\mathcal{N}(x; \mu, \nu) = \frac{1}{\sqrt{2\pi\nu}} \exp\left[-\frac{1}{2\nu}(x - \mu)^2\right]. \quad (9)$$

Using this noise distribution and the mixing model, the expression of the likelihood is obtained as

$$p(\mathbf{D}|\mathbf{S}, \mathbf{C}, \boldsymbol{\theta}_1) \propto \prod_{i=1}^m \prod_{k=1}^n \left(\frac{1}{\sigma_i}\right)^n \exp\left[-\frac{1}{2\sigma_i^2}(D_{(i,k)} - [\mathbf{C}\mathbf{S}]_{(i,k)})^2\right]. \quad (10)$$

3.3 Prior Distributions of Pure Spectra and Concentration Profiles

The pure component spectra are considered mutually independent and identically distributed and each pure spectrum $\{S_{(j,k)}, k = 1, \dots, n\}$ is assumed distributed as Gamma distribution of parameters (α_j, β_j) . These parameters are considered constant for each spectrum over all the spectral band but may differ from one pure spectrum to another. Thus, the prior distribution of the pure spectra is

$$p(\mathbf{S}|\boldsymbol{\theta}_2) = \prod_{j=1}^p \prod_{k=1}^n \mathcal{G}(S_{(j,k)}; \alpha_j, \beta_j), \quad (11)$$

where $\mathcal{G}(z; \alpha, \beta)$ represents the Gamma density for the variable z with parameters α and β . The vector $\boldsymbol{\theta}_2 = [\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p]^T$ contains the parameters of all the prior distributions.

The Gamma density is expressed as

$$\mathcal{G}(z; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp[-\beta x] & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases} \quad (12)$$

where $\Gamma(\alpha)$ is the Gamma function. The choice of this distribution is motivated by the fact that:

- it takes into account explicitly the non-negativity since $p(z < 0) = 0$;
- due to its two parameters α and β , this density can be used to fit various distributions and has the Khi-2, exponential distributions as particular cases [46]. Figure 2, illustrates three typical shapes of the Gamma density. Case (a) is well adapted to encode sparsity (concentration of the distribution in the neighborhood of zero), while case (b) corresponds, for example, to the case of a sparse pure spectrum superimposed on a background, typically consisting in the tail of a wide peak. Case (c) shows that the Gamma distribution can also lead to a noninformative positive prior.

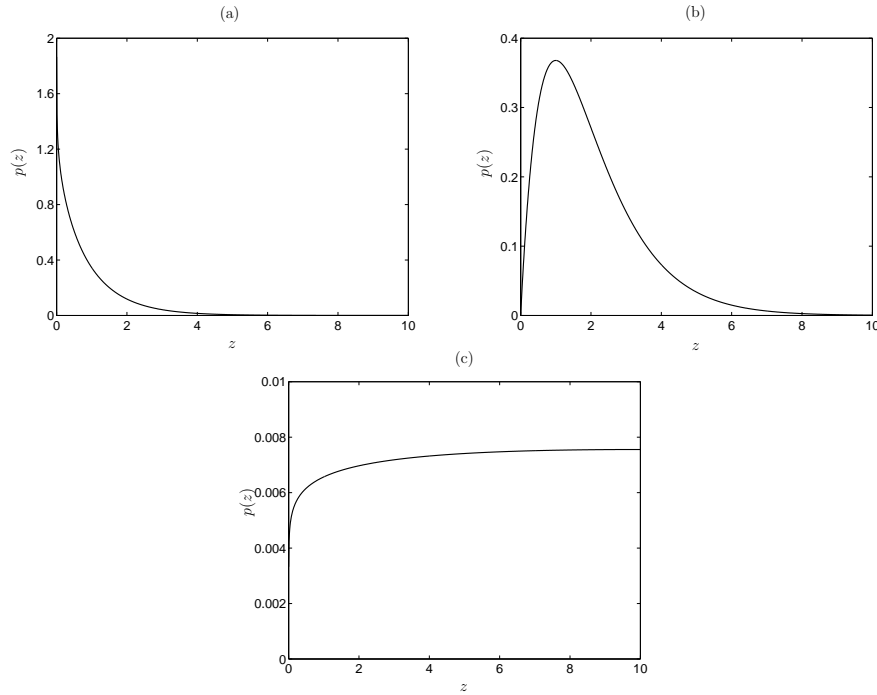


Figure 2. Illustrations of the Gamma distribution shapes according to the values of its two parameters: (a) $\alpha = 0.9, \beta = 1$, (b) $\alpha = 2, \beta = 1$ and (c) $\alpha = 1.1, \beta = 0.1$

Each j -th column of the mixing matrix is also assumed distributed as a Gamma density of parameters (γ_j, δ_j) . These parameters are considered constant for each concentration profile but may differ from one profile to another. The prior distribution of mixing coefficients is then given as

$$p(\mathbf{C}|\boldsymbol{\theta}_3) = \prod_{i=1}^m \prod_{j=1}^p \mathcal{G}(C_{(i,j)}; \gamma_j, \delta_j), \quad (13)$$

where $\boldsymbol{\theta}_3 = [\gamma_1, \dots, \gamma_p, \delta_1, \dots, \delta_p]^T$.

3.4 Posterior Density and Resulting Criterion

Noting $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3]$ and according to equation (4) one can deduce the expression of the posterior law

$$p(\mathbf{S}, \mathbf{C} | \mathbf{D}, \boldsymbol{\theta}) \propto \times \prod_{k=1}^n \prod_{i=1}^m \mathcal{N}(E_{(i,k)}; 0, \sigma_i^2) \times \prod_{k=1}^n \prod_{j=1}^p \mathcal{G}(S_{(j,k)}; \alpha_j, \beta_j) \times \prod_{i=1}^m \prod_{j=1}^p \mathcal{G}(C_{(i,j)}; \gamma_j, \delta_j). \quad (14)$$

The corresponding criterion takes the form

$$J_1(\mathbf{S}, \mathbf{C} | \boldsymbol{\theta}) = Q(\mathbf{S}, \mathbf{C} | \boldsymbol{\theta}_1) + R_S(\mathbf{S} | \boldsymbol{\theta}_2) + R_C(\mathbf{C} | \boldsymbol{\theta}_3), \quad (15)$$

where the terms Q , R_S , and R_C are given by

$$\begin{cases} Q(\mathbf{S}, \mathbf{C} | \boldsymbol{\theta}_1) = \sum_{i=1}^m \sum_{k=1}^n \frac{1}{2\sigma_i^2} (D_{(i,k)} - [\mathbf{CS}]_{(i,k)})^2, \\ R_S(\mathbf{S} | \boldsymbol{\theta}_2) = \sum_{j=1}^p \sum_{k=1}^n (1 - \alpha_j) \log S_{(j,k)} + \sum_{j=1}^p \sum_{k=1}^n \beta_j S_{(j,k)}, \\ R_C(\mathbf{C} | \boldsymbol{\theta}_3) = \sum_{j=1}^p \sum_{i=1}^m (1 - \gamma_j) \log C_{(i,j)} + \sum_{j=1}^p \sum_{i=1}^m \delta_j C_{(i,j)}. \end{cases} \quad (16)$$

The first term Q can be seen as a data fitting measure, while the last two terms are regularization criteria that prevent negative values of \mathbf{C} and \mathbf{S} .

3.5 Bayesian Formulation of the PMF Criterion and Links with the Proposed Approach

In the PMF method [35, 36], the minimized criterion is given by

$$\begin{aligned} J_2(\mathbf{S}, \mathbf{C} | \boldsymbol{\theta}) = & \sum_{i=1}^m \sum_{k=1}^n \frac{1}{2\sigma_{(i,k)}^2} (D_{(i,k)} - [\mathbf{CS}]_{(i,k)})^2 \\ & - \alpha \sum_{j=1}^p \sum_{k=1}^n \log S_{(j,k)} + \beta \sum_{j=1}^p \sum_{k=1}^n S_{(j,k)}^2 - \gamma \sum_{j=1}^p \sum_{i=1}^m \log C_{(i,j)} + \delta \sum_{j=1}^p \sum_{i=1}^m C_{(i,j)}^2. \end{aligned} \quad (17)$$

where $(\alpha, \beta, \gamma, \delta)$ are positive regularization parameters. This regularized least squares method uses on pure spectra and concentrations an *a priori* cost function of the form

$$f(z) = -\alpha \log z + \beta z^2, \quad \text{with } z \geq 0 \quad (18)$$

and corresponds to assign to the variable z a prior distribution defined as

$$p(z | \alpha, \beta) = K(\alpha, \beta) \exp[-f(z)] = K(\alpha, \beta) z^\alpha \exp[-\beta z^2] \quad \text{with } z \geq 0, \quad (19)$$

where $K(\alpha, \beta) = 2\beta^{(\alpha+1)/2}/\Gamma((\alpha+1)/2)$. It can be noted that this distribution corresponds to assuming a Gamma prior distribution on z^2 with parameters $((\alpha+1)/2, \beta)$. Figure 3 shows three typical shapes of this distribution. One can particularly note a similarity with the Gamma distribution profiles shown in Figure 2. To make a comparison between the PMF method and our proposed approach, the following discussion can be pointed out: (i) The case $\alpha = 0$, corresponds to assigning an *a priori* truncated Gaussian distribution. Contrarily to the Gamma distribution, when $\alpha > 0$ this distribution is not appropriate to encode the sparsity since its mode is greater than zero. (ii) In the PMF method, the values of the hyperparameters should be fixed *a priori* and are considered equal for all the pure spectra and for all the concentrations, which is nothing but assuming that all the pure spectra and concentration profiles have the same distribution. In the proposed inference these parameters may differ for different components and due to the probabilistic formulation, all the hyperparameters are estimated jointly with the pure component spectra and concentration profiles. (iii) Due to the non-convexity of the two criteria J_1 and J_2 when the hyperparameter estimation is added, deterministic optimization methods such as gradient or Newton-Raphson algorithms needs either a good starting point or multiple random initializations in order to reach the global minimum. To handle this optimization problem and to avoid a manual choice of the hyperparameters, we propose to use stochastic estimation methods and we include the problem of estimating the hyperparameters in the inference.

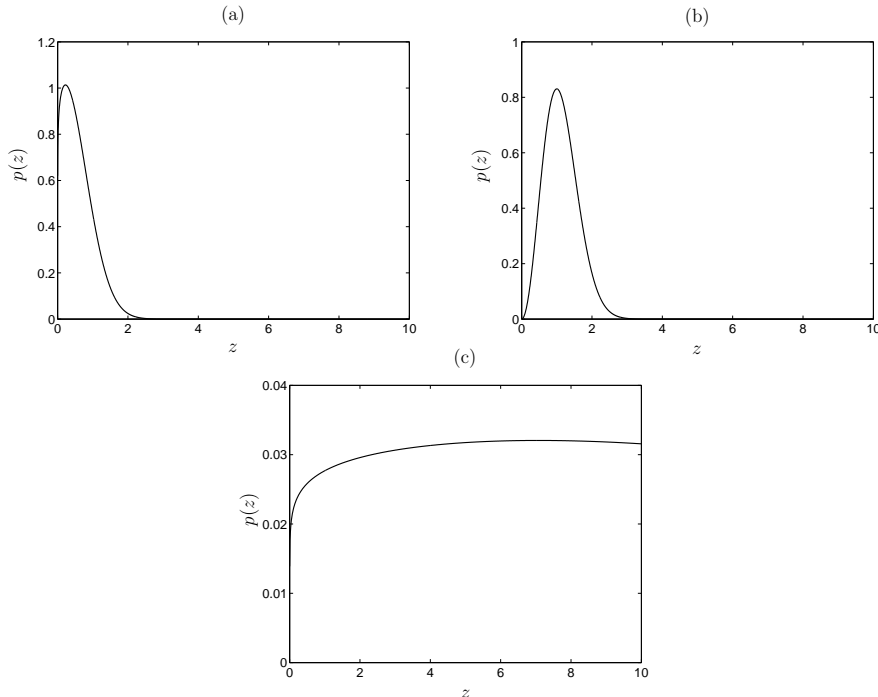


Figure 3. Illustrations of the *a priori* distribution associated to the PMF criterion for (a) $\alpha = 0.1, \beta = 1$, (b) $\alpha = 2, \beta = 1$ and (c) $\alpha = 0.1, \beta = 0.001$.

3.6 Estimation of the hyperparameters

In practice, the parameters of the Gamma distributions as well as the noise variances are not known *a priori*. So, for an unsupervised processing, they must also be considered in the inference. Thus, by using Bayes' theorem and assigning appropriate *a priori* distributions to these hyperparameters, the whole *a posteriori* distribution, including the hyperparameters, is expressed as

$$p(\mathbf{S}, \mathbf{C}, \boldsymbol{\theta} | \mathbf{D}) \propto p(\mathbf{S}, \mathbf{C} | \mathbf{D}, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}). \quad (20)$$

The joint estimation of the pure spectra, the mixing coefficients and the hyperparameters is then performed from this *a posteriori* distribution. However, this inference also needs the definition of appropriate priors for the hyperparameters, that is choosing *a priori* distributions allowing to set the range of admissible values of the hyperparameters, which is much easier than specifying a specific value of each parameter.

3.6.1 Priors on the noise variances

Each noise variance is positive, therefore an appropriate distribution is to assign an *a priori* Gamma distribution to $1/\sigma_i^2$ with parameters $(2, \epsilon)$ where ϵ is set to a small value (10^{-3}). This distribution is chosen because it encodes the positivity and leads to simpler computations of the *a posteriori* distribution $p(\boldsymbol{\theta}_1 | \mathbf{S}, \mathbf{C}, \mathbf{D})$ [45].

3.6.2 Priors on the Gamma distribution parameters

The parameters of the Gamma distributions associated to the pure spectra and concentration profiles are assumed as *a priori* Gamma distributed with parameters $(2, \epsilon)$ where ϵ is set to a small value (10^{-3}). Similarly to the noise variances, this distribution is chosen since it encodes the positivity of the parameters and leads to simpler computations of the *a posteriori* distributions $p(\boldsymbol{\theta}_2 | \mathbf{S})$ and $p(\boldsymbol{\theta}_3 | \mathbf{C})$. Note also that these prior distributions are not very informative (quasi-uniform shape) since their mode is $1/\epsilon$ and their variance is equal to $2/\epsilon^2$.

3.7 Estimation via Markov Chain Monte Carlo Methods

Markov Chain Monte Carlo (MCMC) is a class of stochastic simulation tools for generating random variables from probability distribution functions. These methods are extensively documented in the statistical literature (see the books [47–49] and the references therein). The

main objective of MCMC techniques is to simulate a stationary ergodic Markov chain whose samples asymptotically follow the posterior density $p(\mathbf{S}, \mathbf{C}, \boldsymbol{\theta} | \mathbf{D})$. Estimates of pure component spectra and mixing coefficients are then calculated from the simulated Markov chains. In the sequel, the main steps of the MCMC algorithm for sampling the posterior distribution (20) are given.

3.7.1 Posterior Density Simulation using MCMC

To simulate $p(\mathbf{S}, \mathbf{C}, \boldsymbol{\theta} | \mathbf{D})$, at each new iteration r of the algorithm, the main steps consist in simulating

1. the pure spectra $\mathbf{S}^{(r+1)}$ from

$$p(\mathbf{S} | \mathbf{D}, \mathbf{C}^{(r)}, \boldsymbol{\theta}^{(r)}) \propto p(\mathbf{D} | \mathbf{S}, \mathbf{C}^{(r)}, \boldsymbol{\theta}^{(r)}) \times p(\mathbf{S} | \boldsymbol{\theta}^{(r)});$$

2. the mixing coefficients $\mathbf{C}^{(r+1)}$ from

$$p(\mathbf{C} | \mathbf{D}, \mathbf{S}^{(r+1)}, \boldsymbol{\theta}^{(r)}) \propto p(\mathbf{D} | \mathbf{S}^{(r+1)}, \mathbf{C}, \boldsymbol{\theta}^{(r)}) \times p(\mathbf{C} | \boldsymbol{\theta}^{(r)});$$

3. the noise variances $\boldsymbol{\theta}_1^{(r+1)}$ from

$$p(\boldsymbol{\theta}_1 | \mathbf{D}, \mathbf{S}^{(r+1)}, \mathbf{C}^{(r+1)}) \propto p(\mathbf{D} | \mathbf{S}^{(r+1)}, \mathbf{C}^{(r+1)}, \boldsymbol{\theta}_1) \times p(\boldsymbol{\theta}_1);$$

4. the source hyperparameters $\boldsymbol{\theta}_2^{(r+1)}$ from

$$p(\boldsymbol{\theta}_2 | \mathbf{S}^{(r+1)}) \propto p(\mathbf{S}^{(r+1)} | \boldsymbol{\theta}_2) \times p(\boldsymbol{\theta}_2);$$

5. the mixing coefficient hyperparameters $\boldsymbol{\theta}_3^{(r+1)}$ from

$$p(\boldsymbol{\theta}_3 | \mathbf{C}^{(r+1)}) \propto p(\mathbf{C}^{(r+1)} | \boldsymbol{\theta}_3) \times p(\boldsymbol{\theta}_3).$$

All the stochastic simulation steps including the expressions of the conditional posterior distributions and their simulation techniques are detailed in [44], where this method is termed *Bayesian Positive Source Separation* (BPSS).

3.7.2 Estimation using the Marginal Posterior Mean Estimator

The MCMC sampler is iterated in such a way to have M samples $\{\mathbf{S}^{(r)}, \mathbf{C}^{(r)}, \boldsymbol{\theta}^{(r)}, r = 1, \dots, M\}$ from the posterior distribution (20). The first L samples corresponding to the burn-in of the

Markov chains are discarded and therefore the Monte Carlo approximation for MPM estimation is achieved by

$$\hat{\mathbf{X}} \approx \frac{1}{M - L + 1} \sum_{r=L+1}^M \mathbf{X}^{(r)}, \quad (21)$$

where $\mathbf{X} \in \{\mathbf{S}, \mathbf{C}, \boldsymbol{\theta}\}$.

4 Illustration of the Bayesian Approach using Synthetic Data Sets

4.1 Mixture Data Synthesis

The mixture data are obtained by constructing analytically three spectral signals of $n = 1000$ data points. Each pure spectrum is obtained as the superposition of 15 Gaussian and Lorentzian shapes with randomly simulated location, amplitude and width parameters. The concentration profiles are obtained from a sequential reaction $A + B \rightarrow C \rightarrow D + E$, where the constructed spectra are assumed, respectively, as those of the components A, C, D. The spectra of reactant B and E are supposed non-observable. The evolution profiles are calculated for $m = 10$ measures with randomly chosen constant rates. The retained mixing coefficients are chosen in such a way that there is no zero concentration value in the measured data. Figures 4 and 5 show the simulated synthetic pure component spectra and their concentration profiles.

4.2 Performance Measures and Noise Level Index

The noise level in each i -th mixture is defined as the ratio between the standard deviation of the i -th noise sequence and the root mean square of the i -th noise free mixture. It is expressed in percent. The reconstruction quality of each pure spectrum or mixing coefficient profile is assessed using a measure of dissimilarity between their recovered and their reference values. The dissimilarity is noted *diss* and defined by

$$diss(z_j, \hat{z}_j) = \sqrt{1 - corr(z_j, \hat{z}_j)^2} \geq 0, \quad (22)$$

where $z \in \{S, C\}$ and $corr(z_j, \hat{z}_j)$ is the correlation coefficient between z_j and its estimate \hat{z}_j . The overall estimation performance is assessed using the *global dissimilarity*, defined by

$$gdiss = \frac{1}{p} \sum_{j=1}^p diss(z_j, \hat{z}_j). \quad (23)$$

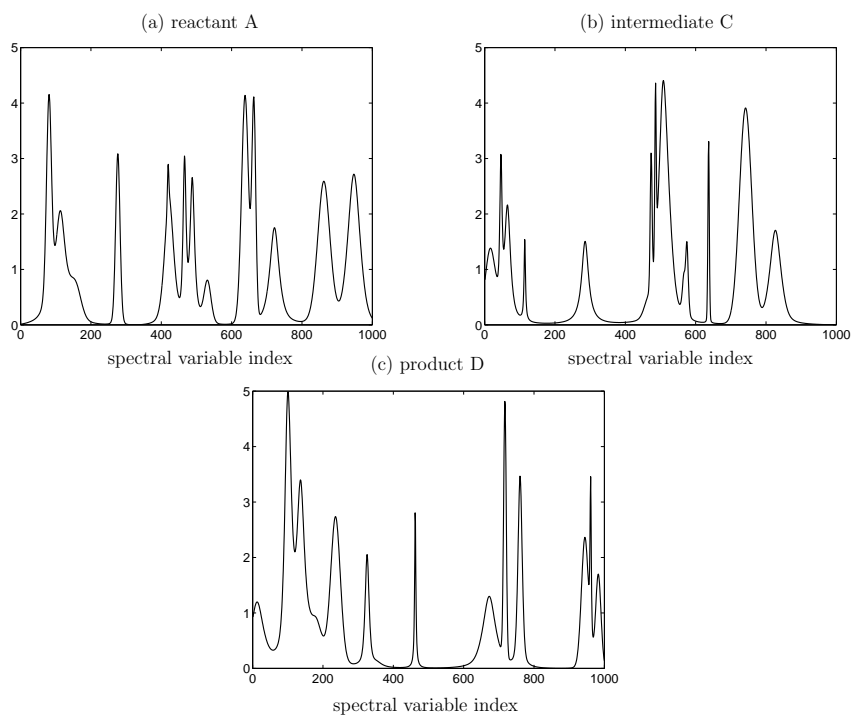


Figure 4. Simulated pure spectra.

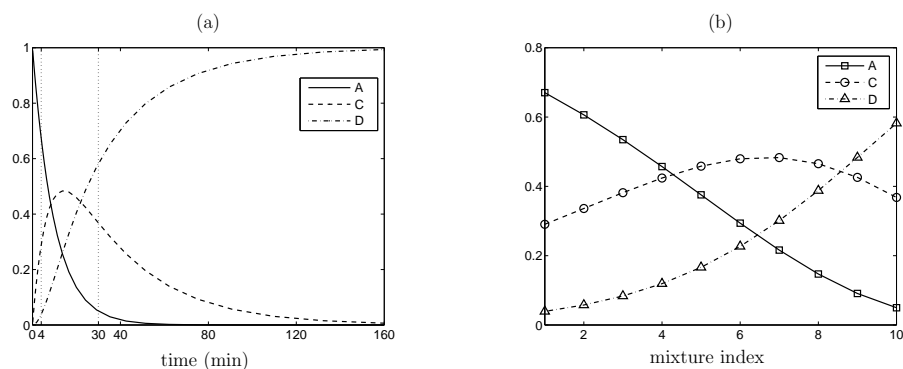


Figure 5. (a) Simulated concentration profiles during all the reaction time and (b) retained mixing coefficients for an observation time interval $[4, 30 \text{ min}]$ with $m = 10$ mixtures.

The estimation is better for lower values of these performance indexes. The results are also evaluated qualitatively by a visual comparison of the reference and recovered spectra and mixing coefficients. Note that the use of the dissimilarity measure is possible only when the reference spectra are available, in an evaluation framework of the methods. Otherwise, a criterion such as the lack of fit can be used to assess how the estimated pure spectra and mixing coefficients can reproduce the observed data. However, it does not measure how well the spectra and concentration profiles are estimated.

4.3 Admissible solutions

A key point in mixture analysis is rotational ambiguity since the resolution using only non-negativity leads to a set of admissible solutions, which in the noise free case are a linear transformation of the true solution [1,11]. A challenging task is to state the admissible solutions and to define further information or assumptions allowing to get a solution where the estimated pure spectra and mixing coefficients are as close as possible to the actual ones. To give an idea about the range of admissible solutions of the synthetic mixture data presented in 4.1, the algorithm of [10] is used to obtain a set of feasible linear transformations of the actual pure spectra and mixing coefficients yielding transformed pure spectra and mixing coefficients which also are non-negative. Assume pure spectra summing to unity and consider the transformation matrix given as

$$\mathbf{T} = \begin{bmatrix} (1 - t_1 - t_2) & t_1 & t_2 \\ t_3 & (1 - t_3 - t_4) & t_4 \\ t_5 & t_6 & (1 - t_5 - t_6) \end{bmatrix}, \quad (24)$$

which ensures to get transformed pure spectra summing to unity thus allowing therefore to handle the scale ambiguity [50]. Figure 6 shows the values of (t_1, t_2) , (t_3, t_4) and (t_5, t_6) yielding a set of non-negativity constrained admissible solutions. It appears clearly that the rotational ambiguity is two-sided since there are transformation matrices with either negative or positive parameters. In that respect, using methods based only on non-negativity, such as ALS or NMF, yield for each initialization one particular solution among these admissible ones.

4.4 Illustration of the method

The proposed algorithm is applied to the mixture data described in section 4.1. Random Gaussian noise is added to the mixture data in such a way to get a noise level of 5% in each mixture. To run the MCMC algorithm, the Markov chains associated to the pure component spectra, the mixing coefficients and the hyperparameters are randomly initialized to random positive values. The sampler is then run for 5000 iterations and the last 1000 simulated samples are retained to calculate the marginal posterior mean estimates. Figure 7 shows the evolution of the marginal posterior mean estimates of the parameters of pure spectra and mixing coefficients *a priori* distributions. Note particularly that the values of these parameters are different for the three components and the values of α_j are less than unity, which corresponds to assigning an *a priori* Gamma distribution of the first case illustrated in figure 2. The final estimates

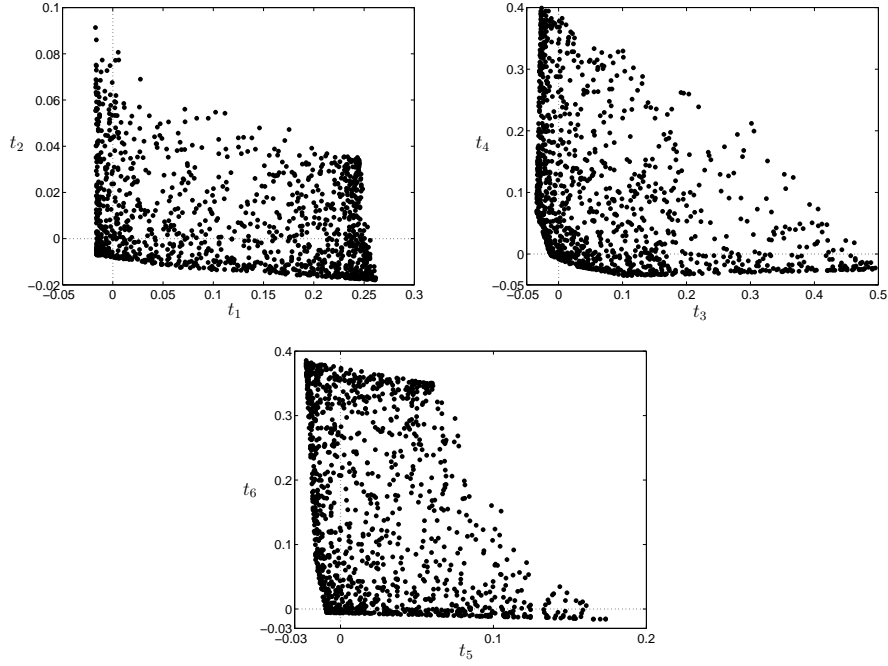


Figure 6. Permitted regions of the rotational parameters. These regions has been obtained using the method of [10]. The algorithm has been run with 1000 random initializations.

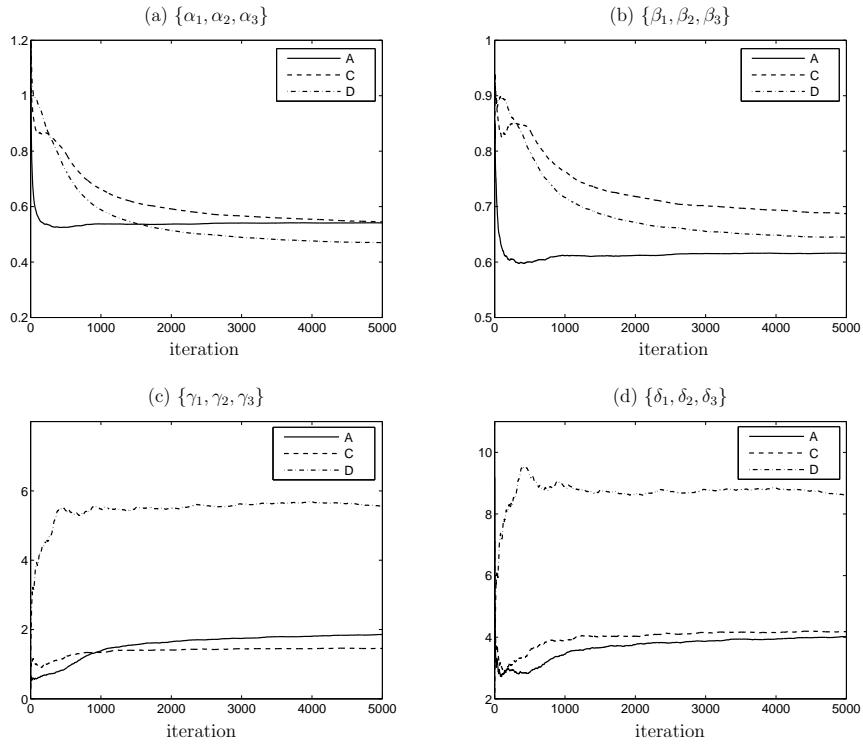


Figure 7. Evolution of the marginal posterior mean estimates of the parameters of the prior distributions associated to the pure spectra and mixing coefficients.

of pure component spectra and mixing coefficient profiles are reported in figure 8. It can be noted the high similarity between the recovered pure spectra and mixing coefficients and their actual values.

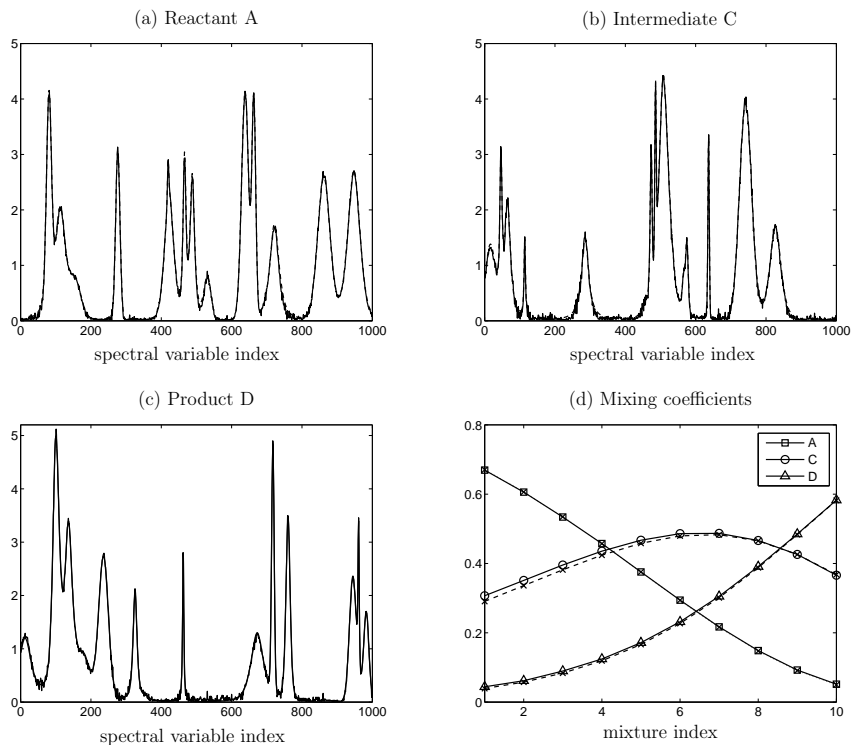


Figure 8. (a–c) Estimated spectra (continuous curves) and comparison with the reference spectra (dotted curves) and (d) estimated mixing coefficients for 5% noise in the mixtures.

4.5 Computation time

The computation time at each iteration of the MCMC algorithm depends on the number of data points, n , in each spectrum, the number of observed mixtures, m , and the number of pure components, p . The overall computation time also depends on the number of iterations necessary to the algorithm convergence. In order to give an idea about the computation time of the algorithm, an experiment is performed by varying the three parameters (m, n, p) and evaluating the time duration of the analysis. Figure 9 shows that the running time of the algorithm increases roughly linearly with all the three parameters (m, n, p). For medium data sets, the analysis can be performed in a reasonable amount of time : for a mixture with dimensions ($m = 30, n = 1000, p = 4$), the running time for 5000 iterations is about 12 minutes. However, for larger data sets and higher number of iterations, the algorithm needs longer time for the analysis.

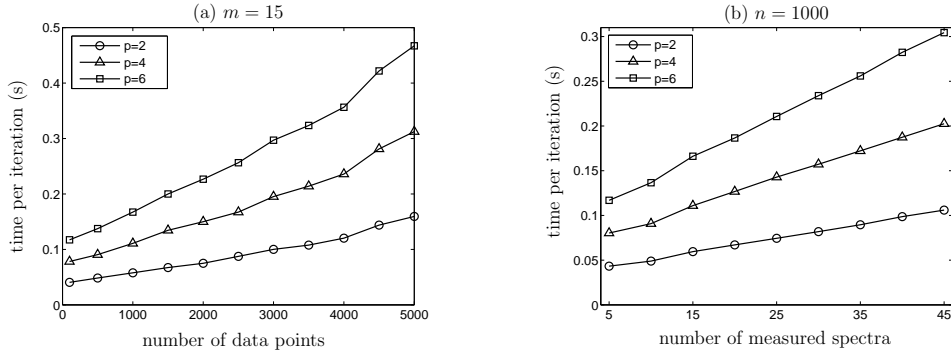


Figure 9. Average computation time per iteration with respect of the data matrix size and number of components.

4.6 Effect of the noise level

The noise level is increased gradually from 1% to 40% and for each value of the noise level the estimation quality of the pure spectra and mixing coefficient is evaluated in terms of global dissimilarity. From figure 10, it can be noted that for low noise levels the pure spectra are reconstructed correctly. For a higher noise level, the pure spectra dissimilarity takes higher values since a noise components remains in the estimated spectra, however the dissimilarity of the mixing coefficients is less sensitive the noise level than the pure spectra. A possible approach to enhance the quality of the pure spectra reconstruction is to include a denoising step during the analysis by introducing a smoothness hypothesis.

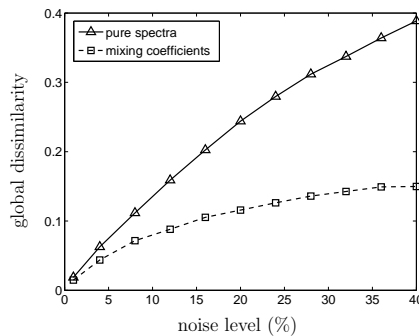


Figure 10. Influence of the noise level on the estimation performances. A Monte Carlo experiment with 100 simulations is performed

4.7 Effect of spectra correlation

In order to discuss the effect of the mutual correlation of the pure spectra on the performances of the BPSS method, a mixture of two synthetic spectra is used. The correlation coefficient of

these spectra is varied while the mixing coefficient are kept fixed. This simulation is performed by a first randomly simulated spectrum which is shifted in order to deduce the second spectrum. Therefore, the correlation coefficient between the two spectra will depend on the value of this shift. The BPSS algorithm is then applied for the analysis of the resulting mixture. The number of algorithm iterations is fixed to 10000 and estimation is performed using the last 1000 realization of the Markov chain. The results of this experience are shown in figure 11, where the performance index of the BPSS method is evaluated with respect to the correlation coefficient of the two spectra and compared with that of the NNICA method. The NNICA method is chosen because this method works under the constraint of orthogonal signals. The two methods reach their best performances as the mutual correlation is small. On the one hand, it can be noted that BPSS algorithm is less sensitive to the mutual correlation of the two spectra as compared to NNICA method. For example, for a correlation coefficient of 0.6 the performance index of the BPSS method remains acceptable. This result can be explained by the fact that, a contrary to the NNICA method, the BPSS algorithm do not optimize an independence measure between the spectra and do not impose strictly their orthogonality.

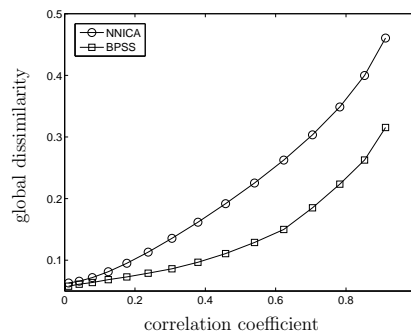


Figure 11. Influence of the spectra correlation on the algorithm performances.

4.8 Effect of a background

To discuss this effect, a mixture of three synthetic spectra where different exponential backgrounds are added to the pure spectra. Figure 12 illustrates the three simulated spectra for which backgrounds with unit amplitude are added. The performances of the analysis are discussed with respect to the background level. The result of this experience is shown in figure 12. When no background is added, the performance index gives an information of the adequation of the Gamma density to represent the pure spectra distribution. By increasing the background level, the pure spectra will be dominated by the exponential background. In this case, the performance index indicates how the Gamma law can represent the background distribution. In the case of two backgrounds, these errors increases leading to a decrease of the performances.

As a conclusion, this simulation shows that for low background amplitude, the Gamma prior model is adequate and leads to good performances but for high background this prior model may be inadequate. In that respect, an extension of the proposed approach is to consider a generalized or shifted Gamma distribution. Another alternative is to include a step of data pre-processing for background removal or to use an *a priori* model in which each pure spectrum may be considered as a superposition of a sparse spectrum with a slowly varying background.

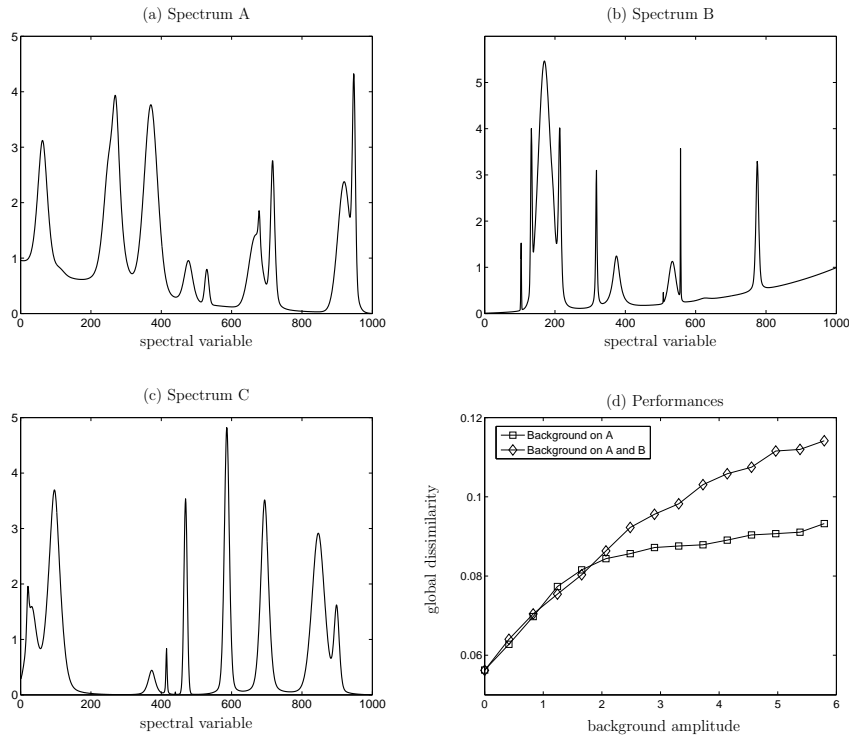


Figure 12. (a,b,c) Pure spectra for a background of unit amplitude in spectra A and B (d) evolution of the performance with respect to the background amplitude.

5 Conclusion

In this paper, a Bayesian approach for spectral mixture data analysis has been presented. The method is based on assigning Gamma distribution as an *a priori* model on pure component spectra and mixing coefficient distributions. This model allows to encode both the sparsity and the non-negativity of these unknown variables. The estimation is achieved using a Markov chain Monte Carlo method which allows to simulate samples from the joint *a posteriori* distribution. The use of this stochastic estimation technique preclude the convergence to local minimums and yields a method working in an unsupervised framework. The only necessary parameters to tune is the number of components and the number of algorithm iterations. Some experiments

with a synthetic data set have been performed to discuss the algorithm efficiency in terms of robustness against the noise level and computation time. It turned out that the method yields effective results even for high noise level in a reasonable time.

A fundamental question in curve resolution is the need to state further information, in addition to non-negativity, allowing to define the best solution among the admissible ones. By incorporating additional assumptions or constraints, like unimodality and closure, the range of admissible solutions will be reduced to a smaller set. On the other hand, a penalized least squares or a Bayesian estimation method leads to one particular solution fulfilling at best the *a priori* distribution model. In the proposed approach, the statistical independence and the sparsity of both the pure component spectra and the mixing coefficients are assumed. Although the Gamma distribution seems to be a good model to encode these information, but nothing guarantee that this prior model is adequate in all circumstances. In that respect, the BPSS method provides pure component spectra and mixing coefficients which are mutually statistically independent and whose statistical distribution is very close to a Gamma distribution. The performances of the BPSS method depend on the ability of the Gamma distribution to represent the actual pure component spectra and mixing coefficient distributions. If the pure component spectra are not Gamma distributed, the solution may be far from the true one and it would be advantageous to use another statistical model. For example, a model allowing to incorporate alternative information, such as smoothness of the component spectra or closure or unimodality of the mixing coefficients profiles.

References

- [1] W. Lawton, E. Sylvestre, Self-modeling curve resolution, *Technometrics* 13 (1971) 617–633.
- [2] W. Windig, Mixture analysis of spectral data by multivariate methods, *Chemometrics and Intelligent Laboratory Systems* 9 (1990) 7–30.
- [3] E. Malinowski, *Factor Analysis in Chemistry*, 3rd Edition, John Willey & Sons, 2002.
- [4] C. Jutten, J. Herault, Separation of sources, Part I, *Signal Processing* 24 (1) (1991) 1–10.
- [5] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution, *Analytica Chimica Acta* 500 (1–2) (2003) 195–210.
- [6] J.-H. Jiang, Y. Liang, Y. Ozaki, Principles and methodologies in self-modeling curve resolution, *Chemometrics and Intelligent Laboratory Systems* 71 (2004) 1–12.

- [7] R. Tauler, A. Izquierdo-Ridorsa, E. Casassas, Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution, *Chemometrics and Intelligent Laboratory Systems* 18 (1993) 293–300.
- [8] R. Tauler, B. Kowalski, S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, *Analytical Chemistry* 65 (1993) 2040–2047.
- [9] N. Ohta, Estimating absorption bands of component dyes by means of principal component analysis, *Analytical Chemistry* 45 (3) (1973) 553–557.
- [10] K. Sasaki, S. Kawata, S. Minami, Constrained nonlinear method for estimating component spectra from multicomponent mixtures, *Applied Optics* 22 (22) (1983) 3599–3606.
- [11] O. Borgen, B. Kowalski, An extension of the multivariate component resolution method to three component, *Analytica Chimica Acta* 174 (1985) 1–26.
- [12] R. Henry, B. Kim, Extension of self-modeling curve resolution to mixtures of more than three components. Part I. finding the basic feasible region, *Chemometrics and Intelligent Laboratory Systems* 8 (1990) 205–216.
- [13] K. Sasaki, S. Kawata, S. Minami, Estimation of component spectral curves from unknown mixture data, *Applied Optics* 23 (12) (1984) 1955–1959.
- [14] S. Kawata, H. Komeda, K. Sasaki, S. Minami, Advanced algorithm for determining component spectra based on principal component analysis, *Applied Spectroscopy* 39 (4) (1985) 610–614.
- [15] Z. Yingzhi, M. Garland, An improved algorithm for estimating pure component spectra in exploratory chemometric studies based on entropy minimization, *Analytica Chimica Acta* 359 (1998) 303–310.
- [16] E. Widjaja, M. Garland, Pure component spectral reconstruction from mixture data using svd, global entropy minimization, and simulated annealing. numerical investigations of admissible objective functions using a synthetic 7-species data set, *Journal of Computational Chemistry* 23 (9) (2003) 911–919.
- [17] P. Comon, Independent component analysis, a new concept?, *Signal Processing*, 36 (3) (1994) 287–314.
- [18] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001.
- [19] A. Cichocki, S. Amari, *Adaptive blind signal and image processing - Learning algorithms and applications*, John Wiley, 2002.
- [20] M. Plumbley, Algorithms for non-negative independent component analysis, *IEEE Transactions on Neural Networks* 14 (3) (2003) 534–543.

- [21] D. Lee, H. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (1999) 788–791.
- [22] C. Lawson, R. Hanson, *Solving Least-Squares Problems*, Prentice-Hall, , 1974.
- [23] R. Bro, S. De Jong, A fast non-negativity constrained least squares algorithm, *Journal of Chemometrics* 11 (1997) 393–401.
- [24] P. Sajda, S. Du, T. Brown, L. Parra, R. Stoyanova, Recovery of constituent spectra in 3D chemical shift imaging using non-negative matrix factorization, in: *Proc. International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2003)*, 2003, pp. 71–76.
- [25] P. Gemperline, A priori estimates of the elution profiles of the components in overlapped liquid chromatography peaks using target factor analysis, *Journal of Chemical Information and Computer Sciences* 24 (1984) 206–212.
- [26] M. Maeder, A. Zilian, Evolving factor analysis, a new multivariate technique in chromatography, *Chemometrics and Intelligent Laboratory Systems* 3 (1988) 205–213.
- [27] F. C. Sanchez, S. Rutan, M. G. Garcia, D. Massart, Resolution of multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis, *Chemometrics and Intelligent Laboratory Systems* 36 (1997) 153–164.
- [28] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Analytical Chemistry* 63 (1991) 1425–1432.
- [29] F. C. Sanchez, J. Toft, B. van den Bogaert, D. L. Massart, Orthogonal projection approach applied to peak purity assessment, *Analytical Chemistry* 68 (1996) 79–85.
- [30] D. Nuzillard, J.-M. Nuzillard, Application of blind source separation to 1-D and 2-D Nuclear Magnetic Resonance spectroscopy, *IEEE Signal Processing Letters* 5 (8) (1998) 209–211.
- [31] D. Nuzillard, S. Bourg, J.-M. Nuzillard, Model-free analysis of mixtures by nmr using blind source separation, *Journal of Magnetic Resonance* 133 (1998) 358–363.
- [32] A. de Juan, Y.V. Heyden, R. Tauler, D. Massart, Assessment of new constraints applied to the alternating least squares method, *Analytica Chimica Acta* 346 (1997) 307–318.
- [33] R. Bro, N. Sidropoulos, Least squares algorithms under unimodality and non-negativity constraints, *Journal of Chemometrics* 12 (1998) 223–247.
- [34] R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, *Journal of Chemometrics* 15 (2000) 627–646.
- [35] P. Paatero, U. Tapper, Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values, *Environmetrics* 5 (1994) 111–126.

- [36] P. Paatero, Least squares formulation of robust non-negative factor analysis, *Chemometrics and Intelligent Laboratory Systems* 37 (1997) 23–35.
- [37] P. O. Hoyer, Non-negative sparse coding, in: *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP'2002)*, 2002, pp. 557–565.
- [38] S. Roberts, Independent component analysis: Source assessment and separation, a Bayesian approach, *IEE Proceedings on Vision, Image and Signal Processing* 145 (3) (1998) 149–154.
- [39] K. Knuth, A Bayesian approach to source separation, in: *Proc. International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aissios, France, 1999, pp. 283–288.
- [40] A. Mohammad-Djafari, A Bayesian approach to source separation, *Proc. International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'1999)*, Boise, Idaho, 1999, pp. 221–244.
- [41] D. Rowe, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*, CRC Press, Boca Raton, Florida, USA, 2003.
- [42] M. F. Ochs, R. S. Stoyanova, F. Arias-Mendoza, T. R. Brown, A new method for spectral decomposition using a bilinear bayesian approach, *Journal of Magnetic Resonance* 137 (1999) 161–176.
- [43] J. Miskin, D. MacKay, Ensemble learning for blind source separation, in: S. Roberts and R. Everson, editors, *Independent Component Analysis: Principles and Practice*, Cambridge University Press, 2001, pp. 209–233.
- [44] S. Moussaoui, D. Brie, A. Mohammad-Djafari, C. Carteret, Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling, to appear in *IEEE Transactions on Signal Processing*.
- [45] C. Robert, *The Bayesian Choice*, 2nd Edition, Springer-Verlag, 2001.
- [46] A. Stuart, J. Ord, *Kendall's Advanced Theory of Statistics*, 6th Edition, Vol. 1: Distribution Theory, John Wiley & Sons, 1994.
- [47] W. Gilks, S. Richardson, D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1999.
- [48] C. Robert, *Monte Carlo Statistical Methods*, Springer-Verlag, 1999.
- [49] M.-H. Chen, Q.-M. Shao, J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York, 2000.
- [50] S. Moussaoui, D. Brie, J. Idier, Non-negative source separation: Range of admissible solutions and conditions for the uniqueness of the solution, in: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2005)*, Philadelphia, USA, 2005, pp. 289–292.